

<https://doi.org/10.21555/top.v760.3239>

# Por el camino de Dennett: una genealogía de la moralidad

## On Dennett's Path: A Genealogy of Morality

Antoni Gomila  
Universitat de les Illes Balears  
España  
toni.gomila@uib.cat  
<https://orcid.org/0000-0001-7325-961X>

Recibido: 03 - 09 - 2024.  
Aceptado: 26 - 05 - 2025.  
Publicado en línea: 20 - 05 - 2026.

Cómo citar este artículo: Gomila, A. (2026). Por el camino de Dennett: una genealogía de la moralidad. *Tópicos, Revista de Filosofía*, 76, 67-85. <https://doi.org/10.21555/top.v760.3239>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

### Resumen

En este artículo discuto los dos capítulos finales de *Darwin's Dangerous Idea* dedicados a la evolución de la moralidad y ofrezco una versión resumida de la genealogía de la moral que he desarrollado en trabajos recientes, como respuesta en parte a las limitaciones señaladas en la propuesta de Dennett. Respecto al capítulo 16 del libro, argumento que se puede ir más allá de su planteamiento y ofrecer una explicación evolutiva del origen y evolución de la moralidad. Respecto al capítulo 17, señalo la inespecificidad moral de su propuesta de un "manual de primeros auxilios morales", lo cual lleva a la necesaria caracterización del ámbito de la moralidad. En la tercera sección se argumenta que la genealogía de la moral que he propuesto permite avanzar en la línea pretendida por Dennet al ofrecer una explicación evolucionista de la moralidad. Esta propuesta no separa la evolución biológica y la evolución cultural como dos procesos alternativos, al modo de Dennet y de la memética, sino que los integra a través del proceso de desarrollo individual, por el cual cada nuevo individuo biológico humano se constituye culturalmente.

*Palabras clave:* moralidad; evolución cultural; Dennett; genealogía de la moral; memética; motivación moral; moralización; perspectiva de segunda persona.

### Abstract

In this paper I discuss the final two chapters of *Darwin's Dangerous Idea* devoted to the evolution of morality and offer a summarized version of the genealogy of morality that I have developed in recent work, in part as a response to the shortcomings identified in Dennett's proposal. Regarding chapter 16 of Dennet's book, I argue that one can go beyond his approach and offer an evolutionary explanation of the origin and evolution of morality. Regarding chapter 17, I point out the moral non-specificity of his proposal for a "moral first aid manual", which leads to the necessary characterization of the scope of morality. In the third section it is argued that this genealogy of morality offers a way to fulfill Dennett's basic goal of an evolutionary account of morality. This proposal does not create a gap between biological evolution and cultural evolution as two alternative processes, as Dennett and memetics assumed, but rather integrates them through the process of individual development, by which each new human biological individual is culturally constituted.

*Keywords:* morality; cultural evolution; Dennett; genealogy of morals; memetics; moral motivation; moralization; second-person perspective.

## 1. Introducción<sup>1</sup>

Coincidió con Dennett en diversas ocasiones. Recuerdo especialmente la primera, en Salamanca, en 1990, en el contexto de uno de los primeros seminarios de filosofía de la mente y ciencia cognitiva, que organizó Fernando Broncano, por la cercanía e interés que mostró con nuestro grupo de jóvenes balbuceantes en inglés. Dennett atendió a todas las presentaciones con interés y respondió a todas las cuestiones que se le plantearon, además de interactuar afable y amistosamente en las instancias informales del encuentro. Mi trabajo se titulaba “Interpretation in Biology and Psychology” (Gomila, 1991), una crítica del instrumentalismo de su “posición intencional” mediante la comparación con el modo en que se atribuyen funciones en biología. En ese momento Dennett estaba acabando la que considero su obra más brillante, *Consciousness Explained* (Dennett, 1992), pero siempre me gustó pensar que mi trabajo tuvo algo que ver con su interés subsiguiente en Darwin y el poder explicativo de la teoría de la selección natural, un movimiento de transición muy significativo en su trayectoria intelectual, iniciada en el marco de referencia del conductismo lógico de Ryle. En cualquier caso, considero a Dennett un modelo ejemplar de cómo hacer filosofía en estos tiempos, desde el naturalismo y la continuidad con la ciencia en diálogo con la cultura en general. Se trata de una filosofía “impura” y desacomplejada, comprometida con la elaboración de los conceptos necesarios para articular nuestra experiencia, y con la pretensión de influir en la elaboración de los conceptos científicos.

La celebración del vigesimoquinto aniversario de la publicación de *Darwin's Dangerous Idea* ofrece una oportunidad excelente para repensar la contribución de Dennett en ese libro pero especialmente para desarrollar su perspectiva. En este artículo quisiera centrarme en la cuestión de la evolución de nuestra psicología moral. Los capítulos finales (16 y 17) de *Darwin's Dangerous Idea* están dedicados a la moralidad y suponen la culminación del esfuerzo desarrollado por Dennett a lo largo de las más de quinientas páginas del libro. El capítulo 18 funciona

---

<sup>1</sup> Este trabajo se inserta en el proyecto “Autonomy as address”, financiado por el Gobierno de la Comunitat València dentro de la convocatoria “Prometeu” (CIPROM/2023/55). . Agradezco a los dos revisores de este trabajo para *Tòpicos, Revista de Filosofia* sus comentarios y sugerencias para mejorar el texto.

como un epílogo en el que se resumen las ideas principales del libro y se anticipan las conclusiones ateas que desarrollará Dennett posteriormente (en relación con la religión, sobre todo). Aquí discutiré los contenidos de estos dos capítulos acerca de la moralidad y de la ética y ofreceré una versión resumida de la explicación evolutiva de la moralidad que, en diversas colaboraciones, he desarrollado en estos últimos años (Pérez y Gomila, 2021, pp. 141-154; Isern-Mas y Gomila, 2022a y 2022b), y que considero que Dennett hubiera visto con simpatía como un modo sólido de desarrollar el programa de una explicación naturalista de la moralidad, tal como él pretendía.

En realidad, lo que llama la atención de estos capítulos es la ausencia de una explicación evolutiva de la moralidad por parte de Dennett, a pesar de que el planteamiento inicial del capítulo 16 remite a tal objetivo. En realidad, ese capítulo consiste principalmente en un crítica a la sociobiología para sugerir, más que detallar, la idea de que nuestra psicología moral es fruto de la evolución cultural, que puede ir en contra de la evolución biológica. Y el capítulo 17 está basado en sus Tanner Lectures de 1988, donde articuló su noción de un “manual de primeros auxilios morales” (Dennett, 1988), propuesta que tiene que ver con la dimensión normativa de la ética. La “peligrosa idea” de Darwin, en relación con la moralidad, para Dennett, supone que la moralidad es el resultado de la evolución, pero de la evolución cultural, que consiste en un proceso mucho más rápido de explorar el espacio de diseño y en encontrar algún “buen truco” que funcione, que no requiere de la herencia genética, y que puede ir contra los intereses genéticos del individuo. Sin embargo, pienso que la caracterización de este “buen truco” deja mucho que desear, del mismo modo que los argumentos que le llevan a esa noción de “manual de primeros auxilios morales” tienen que ver con el razonamiento humano en general más que con el razonamiento y el juicio moral en particular. Ciertamente el tema de la moralidad no aparece en *From Bacteria to Bach and Back* (Dennett, 2017), que puede considerarse la última versión del pensamiento de Dennett sobre la evolución de la mente, lo que puede interpretarse como un signo de que no siguió trabajando sobre este tema.

Nuestro planteamiento, en cambio, parte de una caracterización del ámbito de la moralidad a partir del sentido del deber: de la motivación moral como elemento distintivo. En efecto, lo propio de la moralidad, a diferencia del razonamiento y el juicio, es que el juicio moral es motivador: impulsa a la acción. Si bien no garantiza el cumplimiento

del deber, implica una forma interior de penalización en tal caso, bien la vergüenza o la culpa, según los casos. Nuestra caracterización de la moralidad es por tanto estructural, no de contenido —dado que cualquier cuestión puede ser moralizada, cuestión sobre la que volveremos—. Y permite plantear la cuestión de su evolución de una forma muy clara y operativa. Por ello, creo que constituye un modo riguroso de dar cuenta de la moralidad humana de un modo naturalista, pero no reduccionista, tal como Dennett pretendía.

Así, comenzaremos por revisar el debate de la sociobiología y la posición de Dennett al respecto. En esa sección, mostraremos que su crítica a la sociobiología se combina con una actitud aprobatoria con respecto a la psicología evolucionista, que resulta inconsistente con su posición global sobre Darwin y la evolución cultural, pero que sirve para plantear la necesidad de una explicación evolutiva del origen y evolución de la moralidad que integre, en lugar de escindir, el ámbito biológico con el cultural. La segunda sección señalará la inespecificidad moral de la propuesta de Dennett de un “manual de primeros auxilios morales”, lo cual nos llevará a la necesaria caracterización del ámbito de la moralidad. La tercera sección intentará combinar los resultados de las dos secciones anteriores en una explicación evolucionista de la moralidad, tal como la caracterizamos, que no separa evolución biológica y evolución cultural como dos procesos alternativos, sino que los integra a través de la importancia del desarrollo individual, por el cual cada nuevo individuo biológico humano se constituye culturalmente. De este modo, nuestra propuesta puede considerarse como un modo de superar las limitaciones de lo que Dennett llegó a ofrecer en *Darwin's Dangerous Idea*.

## 2. “On the origin of morality”

El título del capítulo 16 de *Darwin's Dangerous Idea* plantea la cuestión del origen de la moralidad. Se trata de un capítulo extenso, de cuarenta páginas. A pesar del anuncio y de la extensión, Dennett no ofrece en ella una explicación evolutiva de la moralidad. El núcleo del capítulo es su crítica a la explicación evolutiva de la moralidad ofrecida por la sociobiología, por reduccionista, pero no se nos ofrece una explicación alternativa. Sin embargo, el tenor del capítulo parece sugerir que tal explicación alternativa, no reduccionista, es deseable y posible. Y se sugiere que tal explicación debería seguir la línea del planteamiento de la psicología evolucionista, pero con matices.

Antes de la crítica a la sociobiología, Dennett recoge los planteamientos de Hobbes y de Nietzsche. El capítulo abre con una extensa cita del *Leviathan* donde se plantea la analogía de la sociedad con el organismo humano. Para Dennett, Hobbes es el primero que se plantea explicar el origen de la moralidad y la sociedad. En el “estado de naturaleza” hobbesiano ya existen los seres humanos, pero se rigen por relaciones de fuerza y dominación. No existen aún los conceptos normativos: ni justicia, ni bien, ni maldad, ni ley. La vida humana, en tal situación, es solitaria, pobre, brutal y breve, regida por el miedo y la violencia. Hasta que aparece la civilización, la vida en sociedad, gracias a un contrato por el que los individuos, por su propio interés, renuncian a seguir sus deseos y aceptan una autoridad absoluta con poder para limitarlos. Esta autoridad absoluta permite la cooperación al establecer normas y disponer del poder para hacerlas respetar. Dennett menciona los nombres de Rousseau, Locke y Rawls como versiones posteriores, más sofisticadas, de esta “reconstrucción racional” que inaugura la explicación contractualista del origen de la sociedad, y distingue acertadamente el proyecto de una explicación del origen de la civilización del proyecto normativo de la justificación de normas de justicia, que Hobbes funde en uno solo —a diferencia de Rawls, por ejemplo, interesado únicamente en la cuestión normativa—. Pero a Dennett le interesa destacar del planteamiento de Hobbes su afinidad con el evolucionismo: Hobbes articula un punto de partida presocial, donde ya existen agentes capaces de actuar para promover su interés, y un punto de llegada social “evolutivamente estable”, un equilibrio evolutivo. Por eso lo considera un antecedente de su planteamiento adaptacionista, donde las soluciones que funcionan pueden ser descubiertas por la vía de la evolución cultural, si bien Dennett se cuida de añadir que los beneficios actuales de una novedad evolutiva de tal calibre no constituyen por sí mismos una explicación evolutiva. La cuestión radica en explicar el modo en que tales agentes autointeresados pasan a actuar con base en normas en primer lugar, y de qué modo se evitan los aprovechados, los tramposos. Este énfasis en la capacidad de decisión de los agentes humanos es lo que le lleva a rechazar los modelos basados en la teoría de juegos que pretenden ver la sociedad como un organismo compacto, al modo de Hobbes, por reduccionistas. Dennett enfatiza la capacidad de decisión y acción de los agentes humanos, que no está circunscrita por los intereses genéticos. Este aspecto va a

constituir su crítica principal a la sociobiología y su reivindicación de la evolución cultural y la memética.

En la segunda sección, Dennett comenta el proyecto de Nietzsche de una genealogía de la moral, ahora sí bajo la influencia, si bien indirecta, de Darwin. En realidad, sostiene Dennett, Nietzsche reacciona al darwinismo social de Spencer, tan simplista que pretende consagrar y legitimar como normas sociales las fuerzas que Hobbes atribuía al estado de naturaleza. El eje de la genealogía de la moral que propone Nietzsche, por el contrario, consiste en que se da cuenta de que para contratar, es decir, para que el contrato social pueda tener lugar, hacen falta agentes capaces de proyectarse hacia el futuro, de comprometerse, y esta capacidad también debe formar parte del *explanans*. Del mismo modo, los agentes involucrados en el contrato deben ser capaces de reconocer a los demás con quienes interactúan y de recordar interacciones pasadas: llevar la cuenta de a quien le debemos y quien nos debe, así como identificar a quien no cumple. Por ello, Nietzsche propone dos fases en el origen de la moral: en la primera habrían aparecido agentes con tales capacidades de memoria y acción, con una vida social primaria, aún premoral, pero (a diferencia de Hobbes) ya fuera del estado de naturaleza; en la segunda se habría producido la famosa “inversión de los valores” nietzscheana: lo que era bueno en esas formas de vida social premoral (regida por relaciones de dominancia basadas en la fuerza y la ley del más fuerte, que consagra una aristocracia de señores de la guerra) pasó a ser condenado y se promovió la cooperación y la confianza.

Estas secciones parece que van a anticipar la propuesta dennettiana del origen de la moral. En lugar de eso, lo que nos encontramos en las otras dos secciones del capítulo es una extensa crítica de la sociobiología por su reduccionismo. En estas secciones, Dennett critica la visión desarrollada en los capítulos anteriores sobre los memes y la evolución cultural como un proceso independiente y autónomo de la evolución biológica. La mayor parte de la crítica sigue el liderazgo de Philip Kitcher (1993). La sociobiología (Wilson, 1975 y 1978) fue un intento de explicar la conducta social humana mediante los mismos recursos teóricos que la conducta social de los insectos sociales, básicamente por el altruismo de parentesco (genético). Obviamente, tales recursos no permiten dar cuenta de normas morales o del sentimiento de obligación moral, ni de la estructura de la acción humana en general. Además, la sociobiología constituye un ejemplo indiscutible de falacia naturalista, al tratar de derivar el deber ser moral del ser biológico. En sentido positivo, Dennett

simplemente hace referencia a los trabajos de Trivers sobre la evolución de la inversión parental, la selección de parentesco y la reciprocidad, y a los de Axelrod (1984) sobre la estrategia “tit-for-tat” para explicar la evolución de la cooperación. Pero como señaló Darwin mismo en *Descent of Man* (Darwin, 1872a), la conducta de la madre que se sacrifica por una cría no puede considerarse aún dentro de la esfera del *moral sense*, puesto que tal conducta no está guiada por un juicio moral ni por sentido del deber. En este sentido, Darwin sigue resultando una referencia inexcusable para no confundir la evolución de la cooperación o del altruismo (presente en multitud de especies) con la evolución de la moralidad (característica exclusivamente humana). Darwin convierte la explicación de la evolución de la moralidad en la de la emergencia del *moral sense*, la capacidad de juzgar moralmente y sentirse motivado a actuar según tal juicio, siguiendo la idea tradicional de la moralidad humana en la versión específica de Hume.

Más allá, en la dirección de una explicación de la moralidad, Dennett se limita a considerar “what limitations we are born with and what variations there are among us that might have ethical relevance” (1995, p. 481). Pero el desarrollo de este planteamiento es escaso. La línea argumental asume la clásica dicotomía entre biología y cultura y se centra en señalar la dificultad central a la que se enfrenta una explicación biológica de alguna de estas capacidades: que siempre es posible una explicación alternativa en términos culturales —lo cual no excluye, insiste, que alguna explicación biológica pueda ser correcta—. Del mismo modo, de la universalidad de una conducta no puede inferirse que sea genética o instintiva. Pero a nivel propositivo nos encontramos únicamente con una opinión favorable a la psicología evolucionista (Barkow *et al.*, 1993; Pinker, 1997), que propone la selección de módulos de capacidades cognitivas específicas, a los que llaman “algoritmos darwinianos”. Desde esta perspectiva se propone como innata la capacidad de detectar tramposos, por ejemplo. Para este enfoque, la mente humana sería como una navaja suiza compuesta de mecanismos de propósito específico, seleccionados evolutivamente para dar una respuesta adaptativa a problemas de coordinación social.

¿Qué aspecto podría tener en este marco una genealogía de la moral? Si bien Dennett, como ya hemos dicho, no llega a abordar la cuestión, la respuesta debería ser que la psicología moral humana resultaría de una serie de «algoritmos darwinianos», un conjunto de módulos seleccionados por sus consecuencias positivas para afrontar

los problemas de coordinación que tuvieron que afrontar nuestros ancestros. En otros trabajos he argumentado contra un enfoque de este tipo (Gomila, 2014), si bien de modo general, básicamente porque pasa por alto el papel central del desarrollo infantil en la constitución de la mente humana. Dicho de otro modo: la ontogenia no recapitula la filogenia. En primer lugar, porque el contexto en que tuvo lugar la evolución de los homíninos no fue fijo e idéntico en el tiempo; de hecho, la innovación cultural fue modificando el nicho ecológico, y como el desarrollo cognitivo es sensible al contexto, es erróneo pensar que se dio en algún momento del pasado del género *Homo*, un fenotipo cognitivo único y universal. Además, dada la evolución del propio nicho ecológico humano por la intervención de nuestra especie —mediante el uso de herramientas, por ejemplo—, el punto de partida del desarrollo de cada generación va cambiando progresivamente. Esta modificación del proceso de desarrollo ontogenético humano ha sido llamada por Tomasello “efecto trinquete”, para referirse a su naturaleza acumulativa (Tomasello, 1999).

Sin embargo, una de las teorías más influyentes actualmente sobre la psicología moral humana, la de los “fundamentos de la moral” (Haidt y Joseph, 2007), adopta este planteamiento innatista y modularista propio de la psicología evolucionista. Según esta propuesta, venimos equipados de serie con cinco módulos, cada uno de los cuales sería la solución adaptativa a un problema que tuvieron que resolver nuestros ancestros: un módulo para el daño/cuidado, que permite proteger y cuidar a los parientes indefensos y vulnerables; otro relativo a la justicia y la reciprocidad, para permitir la cooperación con los no parientes; un tercero sobre la afinidad y lealtad con el propio grupo, que permitiría la cooperación grupal; un cuarto módulo sobre la autoridad y el respeto, que permite negociar las jerarquías sociales; y un quinto módulo sobre la pureza y la santidad, cuya función sería proteger de los microbios y parásitos. Pero estos módulos iniciales no generan por sí mismos juicios morales, sino que tienen que estructurarse en función de las normas sociales efectivas existentes a través de un proceso de aprendizaje. Y una vez estructurados, funcionan dando lugar a intuiciones basadas en respuestas emocionales, no a razonamientos. No es este el lugar para una discusión en profundidad de esta teoría (para una crítica detallada, cfr. Suhler y Churchland, 2011), pero es importante observar dos cosas: que la evidencia que aporta Haidt tiene que ver con la segunda parte, con el aprendizaje social y la dimensión intuitiva y emocional de los juicios

morales, mientras que para la primera —el innatismo y la modularidad— la evidencia sigue faltando. De hecho, la evidencia disponible apunta en el sentido contrario: que el juicio moral no es modular, sino que es sensible a cualquier aspecto de la situación, por menor o irrelevante que pueda parecer, y novedoso o hipotético (Christensen y Gomila, 2012). En definitiva, la teoría de los fundamentos de la moralidad no da cuenta del origen de las normas sociales, que deben existir ya para que puedan ser aprendidas, ni del sentimiento genuino de obligación moral, que la teoría no contempla. Lo que hace, en realidad, es tratar de simplificar la diversidad moral humana identificando cinco áreas temáticas y enfatizando que los juicios morales son intuitivos, no resultado de largos razonamientos. Pero en realidad puede haber normas morales sobre cualquier cosa, como demuestra el fenómeno de la moralización (Rozin, 1999): fumar se convirtió en moralmente reprochable cuando antes no lo era, la homosexualidad ha dejado de serlo, etc.

Creo que Dennett sería sensible a estas consideraciones, a pesar de la simpatía que manifestó por la psicología evolucionista al enfatizar la evolución cultural y la memética, y la plasticidad humana para desarrollar las conductas más diversas. Por nuestra parte, suscribimos su crítica a la sociobiología pero, como trataremos de mostrar, falta en Dennett una caracterización adecuada de lo que debe ser explicado evolutivamente: la aparición del *moral sense*, del juicio moral motivador de la conducta. Los contenidos de la moral van a ser naturalmente diversos, dependientes del contexto cultural e histórico, pero la cuestión básica es qué convierte las normas en morales en primer lugar. Pero esta caracterización estructural de la moralidad es compatible con una concepción de la evolución cultural como parte de la evolución biológica de nuestra especie: la estrategia adaptativa seguida por nuestro linaje fue precisamente la plasticidad fenotípica que resulta de una mente que necesita tiempo para estabilizarse y se constituye en interacción, con el nicho y con los demás. Y la clave para esta integración es el desarrollo ontogenético. La evolución biológica y la cultural no se contraponen como procesos diferentes e incompatibles, como asume Dennett, sino que se integran a través del proceso de desarrollo mental de cada individuo.

### 3. “Redesigning morality”

En el siguiente capítulo, Dennett trata de ofrecer una ética substantiva de inspiración evolucionista, que llama una “ética naturalizada” —en

alusión al “Epistemology Naturalized” de Quine (1969). En ese capítulo 17 trata de responder a la siguiente pregunta: “¿Qué implicaciones tiene para la ética el hecho de que somos buscadores heurísticos finitos, y con presión temporal, de verdades éticas?” (1995, p. 493). La respuesta a esta pregunta, no obstante, va a tener poco de específicamente moral, a pesar de que se plantee como una cuestión de ética normativa. El argumento que desarrolla Dennett, basado en sus Tanner Lectures (Dennett, 1988), que denomina “Manual de primeros auxilios morales”, consiste en explorar las consecuencias normativas de las limitaciones de la cognición humana en general.

Esto es así porque su adversario dialéctico inicial en este caso es el utilitarismo. El utilitarismo afirma que lo correcto es la alternativa que genera mayor felicidad o placer global. Esta posición se enfrenta al problema de calcular la utilidad, puesto que no es algo sencillo e inmediato. Dennett ilustra su argumento con una analogía. La máxima utilitarista “actúa de modo que maximices la felicidad, o el placer” es como la recomendación “compra barato y vende caro”: no sirve en la práctica para decidir qué comprar y cuándo. Dennett se basa en la crítica de Herbert Simon (1969) a los modelos abstractos que necesitan explorar todo el espacio del problema para poder establecer el razonamiento óptimo. Pero tales cálculos suponen tiempo, esfuerzo y capacidades ilimitadas, elementos todos ellos fuera de nuestro alcance. Como Simon, Dennett propone renunciar a la pretensión de encontrar la opción óptima y conformarse con la que sea satisfactoria. En la práctica, supone substituir la pretensión de disponer de algoritmos —procesos efectivos que garantizan alcanzar la solución óptima— y reconocer que contamos únicamente con heurísticos —procesos rápidos y frugales en términos de procesamiento de información—, que nos ofrecen soluciones satisfactorias. La cuestión subsiguiente de cómo esto afecta al propio proyecto normativo de la ética, no obstante, no se afronta en realidad.

En cualquier caso, Dennett va más allá del rechazo al utilitarismo. Según su razonamiento, lo mismo ocurre con el enfoque kantiano, que dispone actuar según máximas universalizables: ¿cómo decidir si una máxima es universalizable o no?, ¿cómo decidir qué máxima es la adecuada en un contexto determinado si el número de máximas universalizables podría ser infinito? El modo de disimular el problema sería mediante cláusulas *ceteris paribus* e idealizaciones semejantes. El problema de fondo, para ambas teorías éticas que conciben el razonamiento moral como un proceso de deducción normativa, tiene

que ver con el problema de la explosión computacional exponencial que generan.

La solución que propone Dennett, ese “manual de primeros auxilios morales”, sin embargo, se parece poco al modo en que se producen los juicios morales: por intuición, generalmente basada en una respuesta emocional, y como ya señalamos, siendo muy sensibles a los mínimos detalles de la situación. Dennett, en cambio, considera el problema del juicio moral como un problema de las limitaciones del diseño de nuestras mentes como resultado de su origen evolutivo, pero al hacerlo se desliza del plano normativo al psicológico, a cómo implementar nuestra capacidad de encontrar razones para nuestros juicios. Y se limita al ámbito del razonamiento prudencial, de la ponderación de razones a favor y en contra, para el que necesariamente vamos a disponer de un tiempo limitado e información insuficiente para alcanzar juicios garantizadamente correctos. En esa medida, sus consideraciones no son específicamente morales: Dennett aplica al ámbito de la moralidad restricciones propias de la cognición humana en general. De este modo, pasa por alto la especificidad del modo en que se produce el juicio moral —ese sentimiento de obligación que predispone a la acción, así como el modo en que el juicio refleja las normas morales que el agente tiene en consideración—. Estos dos aspectos son los que una genealogía de la moral debe ser capaz de explicar. Y son los que hemos tratado de aclarar en nuestra propuesta.

#### **4. Una genealogía de la moral alternativa**

Nuestra explicación de la evolución de la moralidad humana asume como elemento preliminar la perspectiva de segunda persona de la atribución psicológica: busca en la estructura de la interacción con los demás los aspectos que van a permitir la emergencia de la moralidad en el caso humano. El punto de partida, por tanto, es premoral y se caracteriza por la existencia de agentes capaces de atribuirse estados mentales como modo de organizar sus interacciones, incluyendo las reacciones emocionales. En el libro *Social Cognition and the Second Person in Human Interaction* (Pérez y Gomila, 2021) caracterizamos con detalle la naturaleza de este tipo de atribuciones, que explotan la dimensión expresiva y corporal de la intencionalidad, y en diversos trabajos hemos argumentado que esta forma de atribución está presente en animales no humanos (Pérez-Manrique y Gomila, 2022) y en niños pequeños (Barone y Gomila, 2020). El punto de partida de nuestra genealogía, por

tanto, consiste en una situación donde agentes no morales interactúan y reaccionan emocionalmente a sus interacciones a partir de atribuciones psicológicas básicas que toman como base la dimensión expresiva, corporal, de tales estados, de modo que pueden ser reconocidos directamente.

El *explanandum*, por su parte, como hemos visto en la discusión de Dennett, consiste en explicar el origen y la evolución de nuestra psicología moral, tanto de la motivación moral, del sentirse moralmente obligado, como de la sensibilidad a las demandas de los demás, en base a juicios que implican normas y valores. En otros términos, una explicación evolutiva que se pretenda satisfactoria debe dar cuenta de nuestra psicología moral completa, no de un aspecto particular —la empatía o la prosocialidad, por ejemplo—. Ciertamente puede resultar contenciosa la caracterización que proponemos, pero a su favor debe tenerse en cuenta que es un aspecto central de la comprensión de la moralidad (de Aristóteles a Kant, por lo menos). Ser moral consiste en ser sensible a las razones de orden moral y ello abarca tanto poder juzgar con base en normas morales, más allá de las propias inclinaciones y preferencias, y que tales juicios promuevan la disposición a actuar con base en tales juicios de un modo genuino, lo que se manifiesta en el resultado si no se actúa como se considera moralmente correcto: una sanción en forma de emoción, de culpa o vergüenza. Esta dimensión está ausente de la acción prudencial o estratégica, donde lo que prima es el interés propio. En este sentido, nuestra propuesta va más allá de las explicaciones que confunden la evolución de la moralidad con la evolución del altruismo. La conducta altruista, entendida como aquella en que un agente actúa contra su interés para beneficiar a otro, puede explicarse mediante la teoría de la selección de parentesco, o de la reciprocidad, directa e indirecta, pero no constituye un caso de conducta moral, tal como ya Darwin afirmó en *Descent of Man* en relación con el ejemplo de una hembra que se sacrifica para salvar a sus crías, tal como hemos señalado *supra*.

Por su parte, el *explanans* de una capacidad de esta complejidad no puede ser directo e inmediato: la explicación evolutiva de la aparición de esta estructura agentiva requiere de al menos dos fases. En la primera, explicamos cómo las reacciones emocionales que median nuestras interacciones llegan a depender de normas o valores propios del grupo. En la segunda, explicamos el sentimiento de obligación moral como resultado de la internalización de sanciones inicialmente externas,

sociales. Conjuntamente explicamos la estructura psicológica de la moralidad como universal y derivada de la manera en que interactuamos con los demás, y su contenido como resultado de la interiorización de las normas del grupo en donde se produce el desarrollo.

Con respecto a la primera fase, el núcleo de la propuesta parte de reconocer que nuestras reacciones emocionales constituyen elementos centrales para la regulación y la coordinación social. De nuevo, Darwin es referencia obligada en este punto: en *The Expression of Emotions in Animals and Man* (Darwin, 1872b) argumenta que la expresión de las emociones no es algo secundario y opcional. Al contrario: la función de las emociones es la coordinación y para ello tienen que ser expresadas, para poder ser reconocidas; correlativamente, cada especie es capaz de reconocer fácilmente esa expresión. Además, las emociones con que nos relacionamos con los demás implican un momento de valoración (generalmente denominado *appraisal* en la teorías de las emociones). No se trata aún de una valoración normativa, basada en una norma, pero puede llegar a convertirse en norma si es reconocida y aceptada por los demás, hasta el punto de generalizarse al grupo y convertirse en la base de las expectativas recíprocas de interacción. La idea que se propone, desde la perspectiva de segunda persona, es que las “actitudes reactivas” (Strawson, 1962) que median nuestras interacciones no tienen garantizada su validez inmediatamente. Podríamos recurrir al ejemplo de Darwall (2006) en su caracterización de esta dimensión intrínsecamente interpersonal de la moralidad: el de mi enfado si alguien me pisa; esta respuesta emocional, que implica una valoración de que pisar está mal, se plantea como una demanda de reconocimiento que debe ser validada por la otra parte (en este caso, el pisador). Su reacción puede ser distinta: podría rechazar mi demanda. Por ejemplo, si patino y me caigo, a un espectador cercano podría parecerle divertido y reírse. Esa risa no me haría ninguna gracia y puedo dirigir una demanda parecida hacia quien se burla de mí, sin que tenga ningún efecto, sin que sea reconocida.

¿De qué va a depender que sea reconocida y, por tanto, validada? Proponemos dos aspectos que contribuyen a esta transición: en primer lugar, el intercambio de posiciones entre los agentes que interactúan repetidamente; en segundo lugar, la perspectiva del observador (*bystander*), que puede intervenir para avalar la demanda o para denegar su validez. Cuando se considera la perspectiva de la dinámica de un grupo con esta estructura emergen como “atractores”, como puntos de equilibrio para la coordinación del grupo, ciertas valoraciones aceptadas

por el grupo (por supuesto, pueden emerger también expectativas relativas a diádas o subgrupos). Tales valoraciones equivalen a normas sociales entendidas como expectativas compartidas (Bicchieri, 2006 y 2016). Estas normas no son leyes, no son reglas explícitas, no forman parte de un código público. Son expectativas que regulan la conducta al permitir anticipar las reacciones validadas socialmente, basadas en formas de control social implícitas. No es preciso dedicar mucho esfuerzo a argumentar la mejora de la aptitud biológica de un grupo de este estilo: mayor coordinación, mayores oportunidades de supervivencia.<sup>2</sup>

En la segunda fase, estas normas sociales se convierten en morales al dejar de ser prudenciales y condicionales respecto de las circunstancias del contexto e internalizarse; con ello, obtienen su poder motivador a través del sentimiento de obligación. Dicho de otro modo, caracterizamos el ámbito moral por su fuerza motivacional incondicional, frente a las normas sociales que son contextuales y por tanto condicionales. Es este carácter dual de tales normas, que se experimentan como externas y objetivas tanto como subjetivas (Stanford, 2018), lo que caracteriza el dominio de la moral —no su contenido, que puede ser variado y diverso—.

¿Cómo pudo tener lugar esta transición? Según nuestra propuesta, el sentimiento de obligación moral emergió al producirse una sutil transformación de un cierto tipo de emociones, aquellas que suponen un *appraisal* del propio agente, bien por uno mismo o por el agente con el que se actúa. Estas emociones —orgullo, autoestima, envidia, agradecimiento, culpa, vergüenza, resentimiento...— se denominan a veces “autoconscientes” porque su dimensión valorativa se refiere al propio agente. Como ya observara Darwin (1872b), estas emociones se elicitan por lo que uno piensa que los demás piensan de uno. En esa medida, la valoración tiene también una dimensión de recompensa

---

<sup>2</sup> Hemos ilustrado esta dinámica con el caso de la confianza, que es una actitud compleja que encierra componentes cognitivos y emocionales, y que ilustra muy claramente la dimensión valorativa implícita en la confianza cuando alguien la defrauda. Lo interesante es darse cuenta de que las relaciones de confianza entre agentes individuales gracias a una experiencia previa positiva acaban repercutiendo en el nivel de confianza de un individuo hacia un desconocido, de modo que se incrementa muy significativamente el grado de cooperación dentro del grupo en general. Cfr. Acedo-Carmona y Gomila (2014 y 2019).

o castigo interiorizado: el agradecimiento que percibo me resulta gratificante; el resentimiento, en cambio, me hace sentir culpable, que es una forma de malestar. Además, tales emociones tienen una dimensión motivacional: tu resentimiento por lo que te hice (o omití hacer), me motiva a tratar de reparar la relación (Sznycer *et al.*, 2016).

Es importante en este punto mostrar la ventaja evolutiva de un grupo social con normas morales además de las sociales, frente a un grupo únicamente con normas sociales. Es decir, normas cuya fuerza normativa se ha interiorizado a través del sentimiento de obligación moral, que lleva aparejada también una sanción emocional en forma de vergüenza o culpa en caso de transgresión. La respuesta consiste en reconocer la función de compromiso que supone sentir una obligación (Frank, 1988), que previene cursos de acción miopes. En grupos sociales pequeños, la cooperación puede sostenerse gracias a vínculos afectivos (del tipo de la empatía), pero en grupos más grandes, donde la cooperación depende de la coordinación entre individuos no emparentados, un mecanismo que permitiera un grado de compromiso semejante resultó adaptativo. Y, recíprocamente, disponer de un sistema para extender el cumplimiento normativo incondicional permitió a su vez grupos más y más grandes.

En resumen, la propuesta parte de agentes capaces de interactuar emocionalmente, con un repertorio emocional limitado y estructurado motivacionalmente. Explica, por un lado, la emergencia de normas sociales y, por otro lado, la transformación de un cierto tipo de emociones especializadas en las interacciones intersubjetivas. El resultado de ambos cambios es la emergencia de normas morales como un subconjunto de las normas sociales, aquellas que se vuelven incondicionales al asociarse a un sentimiento de obligación. Este planteamiento permite dar cuenta, además, de la diversidad de normas morales y de los cambios que pueden experimentar, a veces en el curso de una generación, proceso que se denomina “moralización” (Rozin, 1999). Rozin se centra en la moralización del consumo de tabaco, en el sentido de constatar cómo lo que inicialmente se planteaba en términos de conveniencia pasó a plantearse en términos morales. Pero cambios semejantes se observan en procesos de toma de decisión política, con la satanización de adversarios (Bernstein y Gomila, 2024).

A modo de resumen, cabría decir que tomamos de Dennett el reto de ofrecer una explicación evolutiva de la moralidad humana, y que hemos tratado de resolverlo superando las limitaciones que señalamos en los dos capítulos finales de *Darwin's Dangerous Idea*: una caracterización

insuficiente de nuestra psicología moral, y un mayor énfasis en la evolución cultural como proceso desconectado de la evolución biológica. Hemos argumentado que el núcleo de nuestra psicología moral no es ningún contenido, ninguna norma acerca de algún ámbito específico de la vida humana, sino la estructura motivacional específica de los juicios morales. Esta estructura psicológica es la que heredamos evolutivamente y que en el curso de la ontogenia se llena de contenido, de las normas especificadas de cada tiempo y lugar, de cada sociedad humana. Compartimos el rechazo de Dennett de las teorías morales puramente normativistas, que conciben el juicio moral como un proceso algorítmico, pero su alternativa en términos de heurísticas pierde de vista la especificidad moral.

A pesar de todo, y como conclusión, me gustaría insistir en la idea de que Dennett nos proporcionó un modelo de cómo hacer una filosofía naturalista que resultara científica y culturalmente relevante. Y en relación con Darwin y la teoría de la evolución por selección natural, contribuyó a reivindicar la importancia de la perspectiva evolutiva para las ciencias sociales. Y lo hizo siempre con ingenio, con una habilidad especial para el uso de analogías, y con un estilo ameno y entretenido, alejado de la aridez del análisis conceptual. Un camino a seguir.

## Referencias

- Acedo-Carmona, C. y Gomila, A. (2014). Personal Trust Increases Cooperation beyond General Trust. *PLoS One*, 9(8), e105559. <https://doi.org/10.1371/journal.pone.0105559>
- Acedo-Carmona, C. y Gomila, A. (2019). Personal Trust Extends Cooperation beyond Trustees: A Mexican Study. *International Journal of Psychology*, 54(5), 687-704. <https://doi.org/10.1002/ijop.12500>
- Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books.
- Barkow, J., Tooby, J. y Cosmides, L. (1993). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press. <https://doi.org/10.1093/oso/9780195060232.001.0001>
- Barone P. y Gomila A. (2020). Infants' Performance in the Indirect False Belief Tasks: A Second-Person Interpretation. *WIRE Cognitive Science*, 12(3), e1551. <https://doi.org/10.1002/wcs.1551>
- Bernstein, A. y Gomila, A. (2024). The Truth in Social Media. *Topoi*, 44, 127-138. <https://doi.org/10.1007/s11245-024-10039-6>

- Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511616037>
- Bicchieri, C. (2016). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190622046.001.0001>
- Christensen, J. F. y Gomila, A. (2012). Moral Dilemmas in Cognitive Neuroscience: A Principled Review. *Neuroscience and Biobehavioral Reviews*, 36, 1249-1264. <https://doi.org/10.1016/j.neubiorev.2012.02.008>
- Darwall, S. (2006). *The Second-Person Standpoint*. Harvard University Press.
- Darwin, C. (1872a). *The Descent of Man, and Selection in Relation to Sex*. John Murray.
- Darwin, C. (1872b). *The Expression of Emotions in Man and Animals*. John Murray.
- Dennett, D. C. (1988). The Moral First Aid Manual. En S. M. McMurrin (ed.), *Tanner Lectures on Human Values*. VIII (pp. 120-147). University of Utah Press.
- Dennett, D. C. (1992). *Consciousness Explained*. Little, Brown & Co.
- Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster.
- Dennett, D. C. (2017). *Fron Bacteria to Bach and Back*. Penguin.
- Frank, R. H. (1988). *Passions within Reason: The Strategic Role of the Emotions*. Norton.
- Gomila, A. (1991). Interpretation in Biology and Psychology. *Epistemología*, 14, 275-292.
- Gomila, A. (2014). Sobre la relación entre ontogenia y filogenia: una crítica de la Psicología Evolucionista. En S. Español (ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* (pp. 261-283). Paidós.
- Haidt, J. y Joseph, C. (2007). The Moral Mind: How 5 Sets of Innate Moral Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules. En P. Carruthers, S. Laurence y S. Stich (eds.), *The Innate Mind*. 3 (pp. 367-392). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195332834.003.0019>
- Isern-Mas, C. y Gomila, A. (2022a). A Second-Personal Approach to the Evolution of Moral Motivation and the Feeling of Obligation. *Biological Theory*, 17(3), 199-209. <https://doi.org/10.1007/s13752-022-00397-y>

- Isern-Mas, C. y Gomila, A. (2022b). Love, Friendship, and Moral Motivation. *Journal of Theoretical and Philosophical Psychology*, 42(2), 93-107. <https://doi.org/10.1037/teo0000166>
- Kitcher, P. (1993). The Evolution of Human Altruism. *Journal of Philosophy*, 90, 497-516. <https://doi.org/10.2307/2941024>
- Pérez, D. y Gomila, A. (2021). *Social Cognition and the Second Person in Human Interaction*. Routledge. <https://doi.org/10.4324/9781003133155>
- Pérez-Manrique, A. y Gomila, A. (2022). Emotional Contagion in Non-Human Animals: A Review. *WIRE Cognitive Science*, 13(1), e1560. <https://doi.org/10.1002/wcs.1560>
- Pinker, S. (1997). *How the Mind Works*. Norton.
- Quine, W. V. O. (1969). Epistemology Naturalized. En W. V. O. Quine, *Ontological Relativity and Other Essays* (pp. 69-90). Columbia University Press. <https://doi.org/10.7312/quin92204-004>
- Rozin, P. (1999). The Process of Moralization. *Psychological Science*, 10(3), 218-221. <https://doi.org/10.1111/1467-9280.00139>
- Simon, H. (1969). *The Sciences of the Artificial*. MIT Press.
- Stanford, P. K. (2018). The Difference between Ice Cream and Nazis: Moral Externalization and the Evolution of Human Cooperation. *Behavioral and Brain Sciences*, 41, e95. <https://doi.org/10.1017/S0140525X17001911>
- Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy*, 48, 1-25.
- Suhler, C. y Churchland, P. (2011). Can Innate, Modular “Foundations” Explain Morality? Challenges for Haidt’s Moral Foundations Theory. *Journal of Cognitive Neuroscience*, 23(9), 2103-2116. <https://doi.org/10.1162/jocn.2011.21637>
- Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S. y Halperin, E. (2016). Shame Closely Tracks the Threat of Devaluation by Others, Even Across Cultures. *Proceedings of the National Academy of Sciences*, 113(10), 2625-2630. <https://doi.org/10.1073/pnas.1514699113>
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press. <https://doi.org/10.4159/9780674044371>
- Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Harvard University Press.
- Wilson, E. O. (1978). *On Human Nature*. Harvard University Press.

